ELSEVIER

Chemistry & Biology 8 (2001) 521–534

Chemistry & Biology

www.elsevier.com/locate/chembiol

Research Paper

Intracellular protein scaffold-mediated display of random peptide libraries for phenotypic screens in mammalian cells

Beau Peelle, James Lorens, Weiqun Li, Jakob Bogenberger, Donald G. Payan, D.C. Anderson*

Rigel Inc., 240 E. Grand Ave., South San Francisco, CA 94080, USA

Received 28 June 2000; revisions requested 11 August 2000; revisions received 12 February 2001; accepted 9 March 2001
First published online 17 April 2001

Abstract

Background: Mammalian cell screens of peptide libraries for changes in cellular phenotype may identify novel functional peptides and their cognate binding partners, and allow identification of signal transduction network members or proteins important in disease processes.

Results: Green fluorescent protein (GFP) peptide libraries with different structural biases were tested by retroviral expression in A549 carcinoma cells, HUVEC and other cell types. Three different loop replacement libraries, containing 12 or 18 random residues, were compatible with enhanced GFP (EGFP) folding, as was a C-terminally fused random 20-mer library. Library concentrations in A549 cells ranged from ca. 1 to 54 μM. Replacement of loop 3 with known nuclear localization sequence (NLS) peptides, but not with inactive mutants, directed EGFP to the nucleus. Microscopy-based screens of three different libraries for non-uniform localization revealed novel NLS peptides, novel variants of a peroxisomal localization motif, a variety of partial NLS peptides, peptides localized to the nucleolus, and nuclear-excluded peptides.

Conclusions: Peptides can be presented by EGFP in conformations that can functionally interact with cellular constituents in mammalian cells. A phenotypic screen resulting in the discovery of novel localization peptides that were not cell type-specific suggests that this methodology may be applied to other screens in cells derived from diseased organisms, and illustrates the use of intracellular combinatorial peptide chemistry in mammalian cells.

Keywords: Functional screen; Intracellular peptide library; Protein scaffold

1. Introduction

Random peptide libraries expressed in mammalian cells should allow screens for individual peptides which cause a desirable change in cellular phenotype. Peptide library members, presented here in the context of a stable protein scaffold, will likely bind to a cellular macromolecule to cause a measurable change in cell physiology. Such screens do not require prior knowledge of signaling pathways that might be coupled to phenotypic changes, thus novel elements or interactions may be discovered by this process. Affinity-isolated binding partners of such peptides may potentially be drug targets when the phenotypic change in the screened cells is related to processes important to human disease. To examine the composition and function of signaling pathways in yeast, peptide libraries have been screened for specific phenotypic effects [1–3]. The presence of a larger number of expressed cellular proteins and a concomitant increase in signaling pathway complexity may make library screens in mammalian cells more relevant to phenotypic changes associated with human disease processes. Cells contain a variety of cytosolic endo- and exo-peptidases such as the proteasome [4,5], the aminopeptidase bleomycin hydrolase [6], leucine aminopeptidases [7,8], and aspartyl aminopeptidase [9]. Although expression of free small peptides within a cell may be desirable especially when access to a macromolecule's binding site has steric restrictions, these peptides may be subject to catabolism by the above proteases. Small pep-

Abbreviations: EGFP, enhanced green fluorescent protein; FACS, fluorescence-activated cell sorting; GFP, green fluorescent protein; HA tag, influenza hemagglutinin epitope tag; NLS, nuclear localization sequence; PBS, phosphate-buffered saline; PCR, polymerase chain reaction; SAR, structure–activity relationship

* Correspondence: D.C. Anderson;
E-mail: danderson@rigel.com

1074-5521/01/$ – see front matter
PII: S1074-5521(01)00031-X

tides may also be transported out of the cytoplasm into the endoplasmic reticulum by the TAP1 and TAP2 transporters [10]. Therefore, presentation of random peptides using a stable protein scaffold may allow higher levels of peptide to accumulate within the cells by blocking or slowing catabolism or export.

Conformational constraints of peptides may limit proteolytic degradation [11] and protect the library complexity while enriching in levels of conformers less available in unrestrained libraries. By limiting the conformational space of each peptide and thus diminishing the conformational entropy lost upon binding, the affinity of the binding interaction can be increased. Some binding partners may not prefer constrained peptides, however, and might bind instead to more extended conformers [12]. Other peptides may function in α-helical conformations, such as in leucine zipper or coiled-coil peptides [13–15]. Thus, an ensemble of peptide libraries with different conformational biases may be most appropriate for functional screening of signaling pathways.

A good display scaffold should be soluble, non-toxic and well expressed in a variety of cells, should have a detectable readout or activity to allow determination of the folding of individual peptide library members, and should have a known structure, allowing the determination of different sites for the intracellular display of functional peptides with differing structural biases. Libraries with different biases may facilitate interaction with undiscovered binding partners which have individual preferences in their ligand interactions. To allow affinity extraction of peptide library binding partners, the scaffold should not bind tightly to cellular proteins. Protein scaffolds, including thioredoxin [16], green fluorescent protein (GFP) [17], and staphylococcal nuclease [3], have been used for intracellular display in yeast and in *Escherichia coli* [18]. Unscaffolded peptide libraries have been screened in HeLa cells [19].

GFP has a known structure, is stably expressed and non-toxic in a variety of cells, and the retention of fluorescence appears to be a sensitive indicator of the preservation of the native structure [20–22]. Here we characterize in mammalian cells several retrovirally expressed GFP-based peptide libraries with different structural biases. We examine the average intracellular library concentrations, the immunoreactivity of an epitope tag and activity of different nuclear localization sequences (NLSs) placed into a surface loop, and demonstrate that structure–activity data consistent with reported experiments can be obtained when different NLS mutants are put into this surface loop. To test these peptide libraries in a phenotypic screen, we have examined individual library members for non-uniform cellular localization. The results demonstrate that individual phenotypes are readily observable and can depend on the type of library screened, and that novel sequences associated with individual phenotypes can be discovered by this screening methodology.

2. Results

2.1. GFP surface loops can display an hemagglutinin (HA) epitope tag reactive with an anti-HA antibody

To select peptide display sites within GFP, we chose surface loops with the highest temperature factors, which may thus be more flexible. The distances across the base of these loops, 2, 3, and 4, were taken as the distance from the carbonyl carbon of the residue preceding the start of the loop nitrogen of the following residue, and are 3.9 Å, 3.4 Å, and 10.3 Å, respectively. None of the loops are directly in the dimer interface, and results below confirm that GFP does not dimerize tightly enough to occlude access of an anti-HA antibody to the HA tag. To test peptide display in these loops, we individually replaced K156–Q157, D173, and G189–P192 with an HA epitope tag. Jurkat E cells from each infected population were collected, lysed in phosphate-buffered saline (PBS) without detergent, and immunoextracted with anti-HA antibody attached to Sepharose beads. The extract was separated by SDS–PAGE, blotted, and detected with anti-GFP antibody. The whole cell lysate from the same cells was treated in the same fashion. Fig. 1 (top) shows the results of this experiment observed by fluorescence-activated cell sorting (FACS), and observed by immunoextraction/blotting (bottom). Immunoextractions of the HA tag in loops 2, 3 and 4 all gave bands slightly above the molecular weight of wild type enhanced GFP (EGFP) ('WT') as expected. No GFP was extracted with anti-HA antibody for the EGFP control, nor was any GFP extracted from uninfected cells. The rank order of expression based on FACS analysis shown in Fig. 1 is EGFP > loop 3 peptide > loop 2 peptide > loop 4 peptide. A similar order is observed for the three GFP bands observed in the Western blot of the whole cell lysate. These results suggest that EGFP with an HA tag placed into different loops does fold and is fluorescent in a mammalian cell line. Since these species were extracted under non-denaturing conditions, these results show that the epitope tag in the different loops in GFP is displayed on the GFP surface and accessible to anti-HA antibody in cell lysates. In a similar experiment, placement of the first four residues of the HA tag into loops 2, 3 and 4 with the same G_4 linkers, giving a total loop size of 12 residues, gave geometric mean fluorescence values ca. 4-fold higher in each case than the full 11 residue tag (19 residues total, data not shown), suggesting that the level of folded protein may depend on the number of loop-inserted residues.

2.2. Comparison of GFP loop library expression in stably infected cells

To compare the relative levels of folded protein in mammalian cells retrovirally expressing the loop 3 and 4 random 12-mer and 18-mer peptide libraries (3.12, 3.18, 4.12,

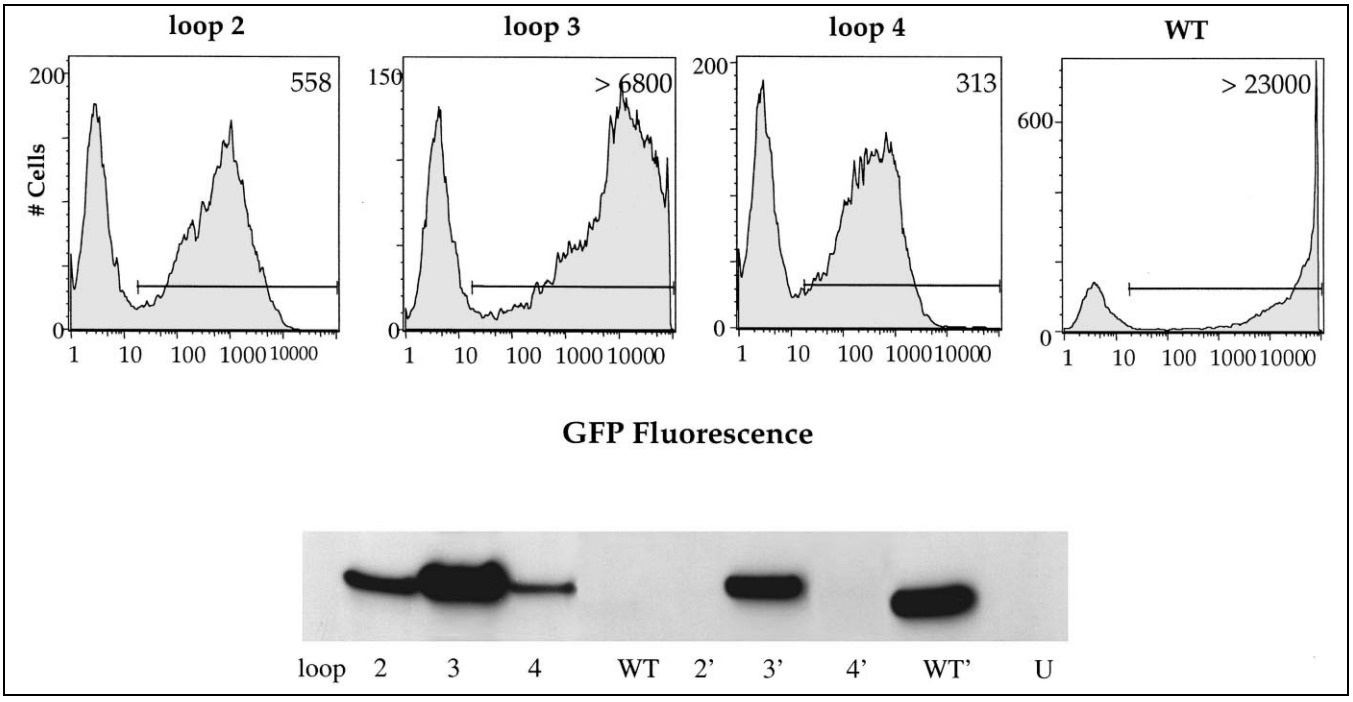


Fig. 1. FACS analysis of an HA epitope tag constrained in loops 2, 3 or 4 of EGFP. Jurkat E cells retrovirally expressing different EGFP loop-constrained sequences were analyzed for GFP expression by FACS (top). The constructs were immunoextracted in native conditions from the lysate of ca. 2×10^4 cells with anti-HA antibody, separated by SDS–PAGE, and Western-blotted using anti-GFP antibody (bottom) shown next to the corresponding whole cell lysates (bottom, primed letters). WT, wild type EGFP; U, uninfected cells; 2, 3, and 4, HA tag constrained in loops 2, 3 and 4. The geometric mean fluorescence is listed in the upper right corner of each FACS plot. In most cases, the level of EGFP or modified EGFP correlates with band intensity observed in the Western blot.

4.18), we analyzed the GFP fluorescence by FACS. Fig. 2 shows the FACS plots of live Jurkat E cells 7 days after infection. The observed geometric mean fluorescence values of GFP expressing cells are listed in the figure. The fluorescence intensities for the above samples relative to EGFP were 0.098:0.056:0.013:0.0083, respectively. Mock infection with the retrovirus parent vector gave the autofluorescent background shown. Thus, the loop 3 libraries fluoresce more intensely than their loop 4 counterparts, while the 12-mer library in loop 3 appears to be somewhat more fluorescent than the 18-mer library.

FACS was also used to compare the relative expression levels of the different libraries in A549 small cell lung carcinoma cells. After 3 days growth post-infection, live cells were examined by FACS (data not shown). Comparing EGFP to a loop 3 random 12-mer and 18-mer library, and to a loop 4 random 12-mer and 18-mer library, the normalized relative geometric mean fluorescence values were 1.0:0.16:0.086:0.012:0.0055, respectively. As with the Jurkat E cells, the loop 3 libraries are expressed at higher levels (ca. 10-fold) than those in loop 4, and the random 12-mers are expressed roughly 2-fold above the random 18-mers. Similar results were obtained by identical methods in BJAB cells, CA46 B cells, 293 T cells, and primary bone marrow-derived murine mast cells (data not shown). When the fluorescence intensity of the loop 3 and 4 random 12-mer and 18-mer libraries was compared with the intensity of staining using a polyclonal anti-GFP antibody coupled to the dye Alexa 568 in A549 cells, the Alexa staining amplitude increased proportionally to the GFP fluorescence intensity (data not shown). This suggests that the GFP fluorescence intensity reflects the amount of immunoreactive GFP protein. The lack of cells with high Alexa 568 fluorescence and low GFP fluorescence suggests that the cellular level of unfolded (non-fluorescent) GFP is low. EGFP and all of the loop 3 and 4 replacement libraries had excitation maxima at ca. 486 nm and emission maxima at 510 nm (data not shown), suggesting there is no net perturbation of the fluorophore and its immediate environment in these loop insert libraries. EGFP levels increased with a noticeable lag after infection, reaching steady state fluorescence levels in ca. 48 h (data not shown). Each of the loop replacement peptide libraries accumulated at slower rates and with more pronounced lag phases than the wild type protein. The lower expression levels of the four peptide libraries seem to correlate with a slower time-course of accumulation of the fluorescent protein. Thus some step prior to the production of the fluorophore in its folded environment may limit the level of the library in the cell.

2.3. Most peptide library sequences allow GFP folding

To examine the folding of different peptides in loops 3 and 4 of EGFP, we sequenced 94 random peptides from each of the four loop libraries. These clones were individ-

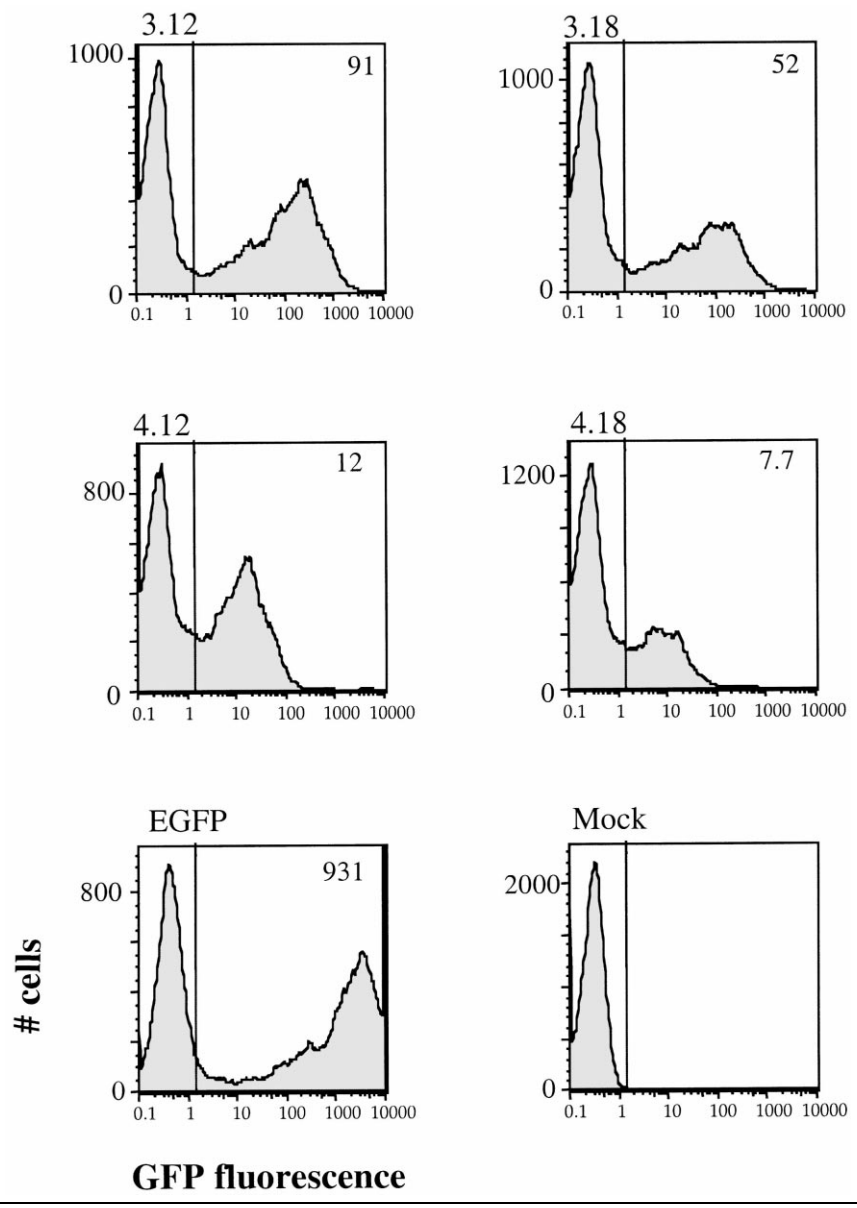


Fig. 2. Jurkat E cell expression of loop 3 and 4 random 12-mer and 18-mer peptide libraries. Jurkat E cells (1×10^5, 7 days post-infection), excited at 488 nm with an argon laser, were analyzed by FACS for GFP fluorescence. The observed geometric mean fluorescence values are listed for each library in the upper right hand corner of the FACS plot. The relative levels for the different libraries (EGFP: 3.12, loop 3 random 12-mer library: 3.18, loop 3 random 18-mer library: 4.12, loop 4 random 12-mer library: 4.18, loop 4 random 18-mer library) are 1.0:0.098:0.056:0.013:0.0083. All libraries are expressed on average at levels ca. 10-fold above background fluorescence or higher.

ually transfected into 293 T cells, and examined for fluorescence using UV-microscopy or FACS. The observed stop codon frequencies were close to the predicted values for the nucleotides coding the 12-mer and 18-mer random peptides. All of the open reading frame sequences without stop codons in the loop 3 libraries gave fluorescent cells. All but one sequence in the loop 4 random 12-mer library resulted in fluorescent cells. This peptide, TTDPPP-GGTQPG, contained four prolines and may thus be constrained to a conformation which is not compatible with formation of a stable loop with a very short base. One of the two 18-mer peptides that did not result in observable GFP fluorescence of the loop 4 library, MLWWV-CTGGGPWFWDQKS, contained four tryptophans, a phenylalanine and a proline, and may not be able to pack into a loop compatible with GFP folding.

2.4. Determination of the intracellular concentration of retrovirally expressed GFP-based peptide libraries

The concentration of individual peptide library members will control the degree of saturation of potential binding partners and, thus, the frequency of obtaining peptides that cause phenotypic changes in the cell. The concentra-

Table 1
NLS sequences constrained in loop 3 of EGFP[a]

Origin	Sequence including linkers	Reported NLS activity
SV-40 NLS	-GQGGG-PPKKKRKVA-GQAGGGGG-	active [25]
SV-40 mutation	-GQGGG-PPK**T**KRKVA-GQAGGGGG-	inactive [25]
SV-40 short linkers	-G-PPKKKRKVA-G-	active
c-myc NLS	-GQGGG-PAAKRVKLD-GQAGGGGG-	active [27]
c-myc mutation	-GQGGG-PAAK**TKR**LD-GQAGGGGG-	inactive [28]
Control flag tag	-GQGGG-VADYKDDDDKAV-GQAGGGGG-	none

[a]Mutations which inactivate the NLS activity of the peptide are shown in bold.

tion of GFP libraries used previously in yeast was not reported. A staphylococcal nuclease loop-insertion/truncation library was present at 4% of cellular protein, or ca. 200 μM [3]. Such libraries are produced using multiple-copy plasmids, and are likely produced at higher levels than libraries produced by retroviral insertion into the host genome. It is thus of interest to examine the magnitude of the difference in expression level in mammalian cells. The lack of detectable perturbation of the average fluorescence excitation (absorption) and emission spectra of individual peptide libraries (data not shown) suggests that the intracellular concentration of each peptide library might be estimated using absorption or fluorescence. A549 cells infected with the TRA vector [38] were sorted for GFP fluorescence, grown for 14 days, and the concentrations of the different libraries were estimated using absorption at 489 nm. EGFP was expressed at 140 μM, the GFP C-terminal random 20-mer library was expressed at 54 μM, the loop 3 random 12-mer library was expressed at 27 μM, and the random 18-mer library was expressed at 19 μM. The loop 4 random libraries were expressed at ca. 20-fold lower levels than the loop 3 libraries. Thus replacement of some of the residues of each loop decreases the expression level in a loop-specific, as well as in a length-dependent, fashion. Placing more residues into either loop correlates with a decreased protein expression.

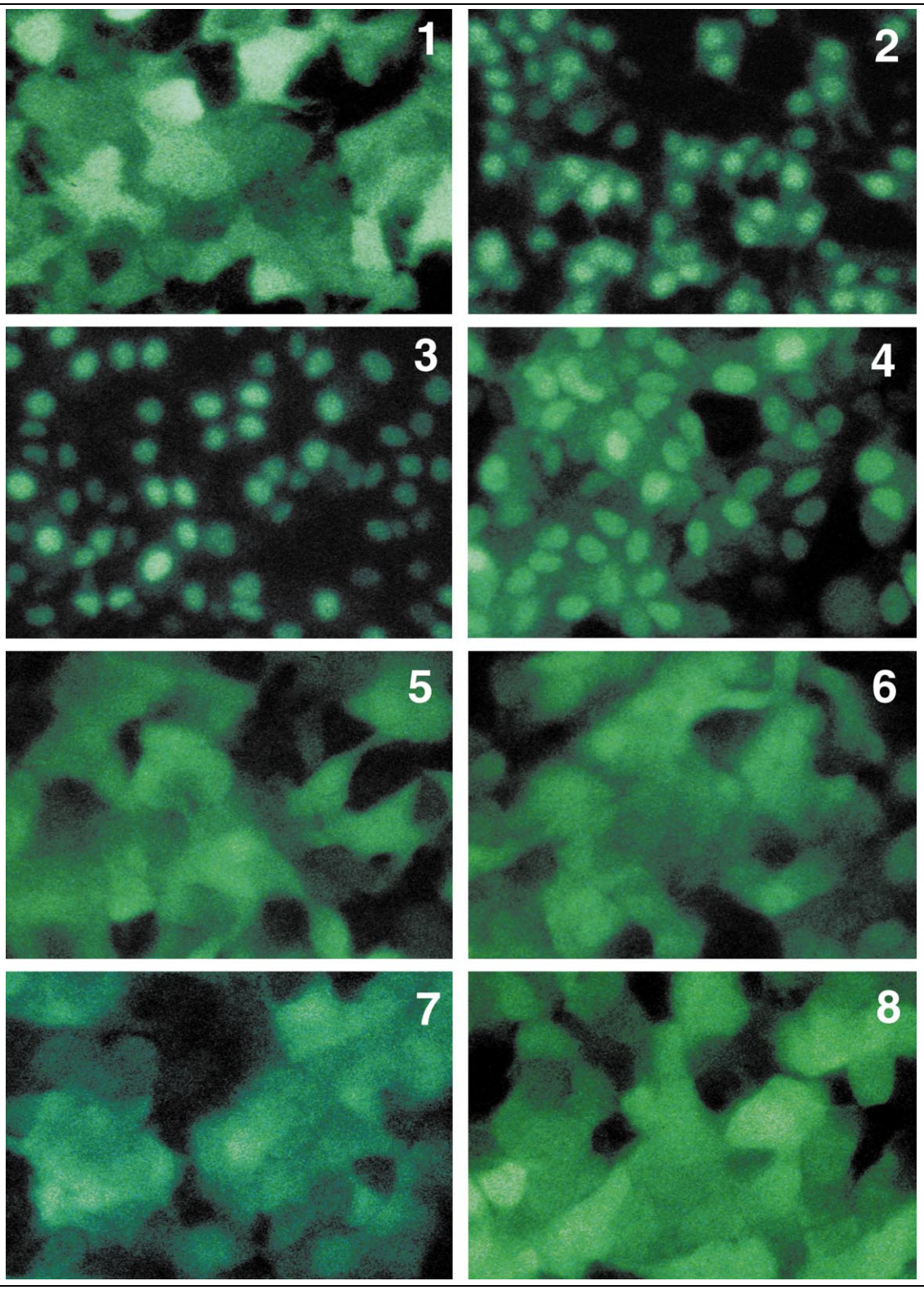
2.5. *The structure–activity relationship (SAR) of NLS peptides inserted in loop 3*

To see if GFP loop replacement peptides can functionally interact with other proteins within the cell, we replaced residues in loop 3 with active and mutationally inactivated NLSs (Table 1). The SV-40-derived NLS interacts with karyopherins in the nuclear pore complex for transport into the nucleus [23,24], is inactivated by mutation of Lys-3 to Thr [25], and binds to importin α in an extended β-strand [26]. The c-myc NLS [27] is inactivated when the central RVK is mutated to TKR [28]. The cellular localization was tested for these sequences by fluorescence microscopy, with the multi-glycine linkers and with a shorter mono-glycine linker.

A549 cells with retrovirally expressed EGFP loop 3 or C-terminal NLS peptides were grown for 5 days and examined by fluorescence microscopy. EGFP (Fig. 3, panel 1, no NLS), EGFP with a loop 3 inactive SV-40 NLS mutant (panel 5), EGFP with a loop 3 inactive c-myc NLS mutant (panel 6), and EGFP with a flag epitope tag in loop 3 (panel 8) showed no nuclear localization. EGFP with an SV-40 NLS added to its C-terminus (panel 2), with an active SV-40 NLS in loop 3 (panel 3), or with an active c-myc NLS in loop 3 (panel 4) showed nuclear localization of EGFP. Thus the NLS peptides in loop 3 or fused to the C-terminus, that are expected to be functional, do in fact localize EGFP to the nucleus. NLS mutants expected to be inactive are in fact inactive, resulting in EGFP fluorescence throughout the cell. When the SV-40 NLS is flanked in loop 3 with only single-glycine linkers (panel 7), no nuclear localization is observed, showing that having only a short linker sequence can compromise the presentation/function of this peptide. These results suggest that loop 3-constrained NLS peptides retain their function when presented within the cell by the EGFP scaffold, and that reported inactive mutants remain inactive in the loop-constrained environment.

2.6. *GFP C-terminally fused epitope tags are not significantly proteolysed in a cell*

GFP loop-constrained peptide libraries will be structurally biased to members with both ends constrained at a fixed short distance. Such libraries may contain short stretches of peptide in an extended conformation (vide supra) but will be unlikely to have long extended (linear) peptide conformers. Since the extended conformation can be important for functional peptides [12], and to establish the feasibility of a C-terminal peptide library using GFP as a scaffold, we examined the fusion of different peptides to the C-terminus of EGFP. The linker sequence, EEAAKA, combined with the protruding endogenous C-terminal sequence of EGFP, MDELYK, was designed to project the fused peptides from the β-can structure. Three different epitope tags, flag DYKDDDDK **1**, influenza HA YPYDVPDYASL **2**, and I-kappa B ERPQEWAMEGPRDGL **3**, were used. After expression for 7 days in Jurkat E cells, GFP fused to each epitope tag was observed on a Western blot (detected with the cognate antibody) slightly above EGFP without additional bands or a smeared or diffuse band, suggesting that a significant part of the additional sequence survives intact in the cell (data



not shown). Seven additional sequences fused to GFP-EEAAKA were selected at random and observed by linear mode matrix-assisted laser desorption ionization-time of flight (MALDI-TOF) mass spectrometry after purification of the construct by anti-GFP affinity chromatography: PAEHKGRNQHSGVKNSRNRR **4**, HCRADKGRASLLNSIEASTGVGP **5**, RFSDPNPTDHDNHMRRWAIP **6**, PHKPRNITKDEKAGLSR **7**, CQQHLRGS **8**, NAKRDSTVALEVFAEARGEC **9**, and GRDKITYSRISTVTSKWPMV **10**. All GFP constructs except **4** and **7**

Fig. 3. SAR of active and inactive NLS peptides when inserted in loop 3 of GFP. A549 cells were infected with retrovirus encoding GFP with different NLS peptides (see Table 1 for sequences) and observed by fluorescence microscopy. Panel 1: cells retrovirally expressing EGFP and a control modified EGFP without an NLS are distributed throughout the cell. Panel 2: cells expressing EGFP with a C-terminal SV-40-derived NLS show modified EGFP localization mainly in the nucleus with a faint cytoplasmic background. Panel 3: cells expressing EGFP with a loop 3 SV-40 NLS show strict nuclear localization of the modified EGFP. Thus, this peptide appears to be functional when displayed in loop 3. Panel 4: cells containing EGFP with a loop 3 c-myc-derived NLS show mainly nuclear localization with some cytoplasmic modified EGFP. This peptide also appears to be functional when in loop 3. Panel 5: cells expressing an inactive mutant of the SV-40 NLS are distributed throughout the cell. Thus, this mutant NLS appears to be inactive when displayed in loop 3 of GFP. Panel 6: cells expressing EGFP with an inactive c-myc NLS peptide in loop 3 also show no localization of EGFP, suggesting that this mutant NLS is also inactive when in loop 3. Panel 7: cells expressing the SV-40 active NLS inserted into loop 3 with only mono-Gly linkers on either side. This peptide now appears to be inactive as the modified EGFP shows no specific nuclear localization. The overall fluorescence of this EGFP is lower than in the other panels. Panel 8: cells expressing EGFP with a loop 3 flag epitope tag show a general cellular distribution of EGFP, as expected. The NLS SARs determined in loop 3 are consistent with previously reported solution NLS SAR.
←

were observed at the expected molecular masses. To check the exact cleavage sites, the constructs were cleaved under native conditions by lys-C endoprotease, and the C-terminal peptides identified by comparison of the peptide maps with peptide maps from GFP alone. Using MALDI-reflectron TOF mass spectrometry, peptide **4** had lost both C-terminal arginines while peptide **7** had lost its C-terminal 13 residues and, separately, the C-terminal Arg. However both cleavages were observed when the synthetic analogs were added to cell lysates, suggesting that the cleavage occurred after cell lysis in spite of the prior addition of a cocktail of protease inhibitors. Thus all 10 C-terminal sequences tested appeared to be stable to proteolysis in these cells when fused to the C-terminus of GFP.

2.7. Individual GFP loop 3 random 12-mer, 18-mer, and C-terminal random 20-mer peptide library sequences have a defined cellular localization in HUVEC and A549 cells

To examine the above three GFP libraries in an intracellular screen, we constructed ca. 10^7 member GFP loop 3 random 12-mer, random 18-mer, and C-terminal random 20-mer peptide libraries in a tetracycline-inducible retroviral vector. The C-terminal 20-mer library was introduced into HUVEC cells, and all were introduced into A549 cells by retroviral infection. After growth for 3 days the cells were visually screened by fluorescence microscopy for clones displaying a non-uniform subcellular location of GFP fluorescence. The cells were grown for this time period to allow several rounds of cell division, so that individual non-lethal non-uniform fluorescence phenotypes resulting from individual heritable peptides would show up in several identical cells in a cluster. On average there were four–six cells in each cluster.

Different localizations were described as nuclear, partial nuclear (most fluorescence in the nucleus, but additional fluorescence in the cytoplasm), nucleolar (nuclear fluorescence with more intense fluorescence in the nucleolus), partially or fully nuclear-excluded, and punctate (fluorescence localized to 'dots' within the cell).

Fig. 4 shows the results from infection of HUVEC cells with the C-terminal 20-mer library. Fig. 4A shows the uniform cellular fluorescence of the EGFP scaffold. Fig. 4B shows the punctate perinuclear localization of EGFP with the novel C-terminal sequence ARIHRL, which was truncated by a stop codon after the L. Fig. 4C shows a different punctate perinuclear localization of GFP fused to the peptide SRTASTGWYTSSNLRRKSAL. Co-labeling these cells with transferrin (data not shown) gave a different pattern of labeling than that observed with this EGFP-fused peptide, showing that the labeled structures are most likely not early endosomes. LysoTracker co-labeling of these same cells, shown as red in Fig. 4F, was most visible in uninfected cells, and clearly different from the green labeling (arrow) due to GFP peptide. These punctate structures are thus probably not part of the lysosomal compartment. Fig. 4D shows the nuclear localization of GFP fused to KRMKTAVVRRRPLRDRRALW. In Fig. 4E, GFP fused to CNNTSDGLRTPKKHKRGSTI is completely excluded from the nucleus although it has a KKXK consensus NLS motif [27]. Each peptide sequence was re-introduced into naive cells, resulting in the same localization pattern. All four localization peptides also exhibited the same phenotype in MRC-5 human fibroblast cells (data not shown).

Similar localizations were seen when all three GFP peptide libraries were introduced into A549 cells. Table 2 presents statistics resulting from observation of 800–2000 cell clusters by fluorescence microscopy. GFP alone was uniformly distributed in each of the 2000 cell clusters examined. When GFP was C-terminally fused to random 20-mers from the library, 4.8% of the different GFPs exhibited some localization, with the most prominent being punctate, partial nuclear localization, and complete nuclear localization. A small number of sequences were nuclear-excluded or nucleolar-localized. The loop 3 random 12-mer library exhibited 3.3% overall localization. The most common localization patterns were partial and strong nuclear localization, with lesser numbers of nucleolar-localized, nuclear-excluded and punctate/granular peptides. The loop 3 random 18-mer library had 4.3% of the tested library members exhibiting some localization. The rank order of occurrence of the distinct patterns of localization differed for each peptide library. Notably, the punctate localization is only present in the C-terminal li-

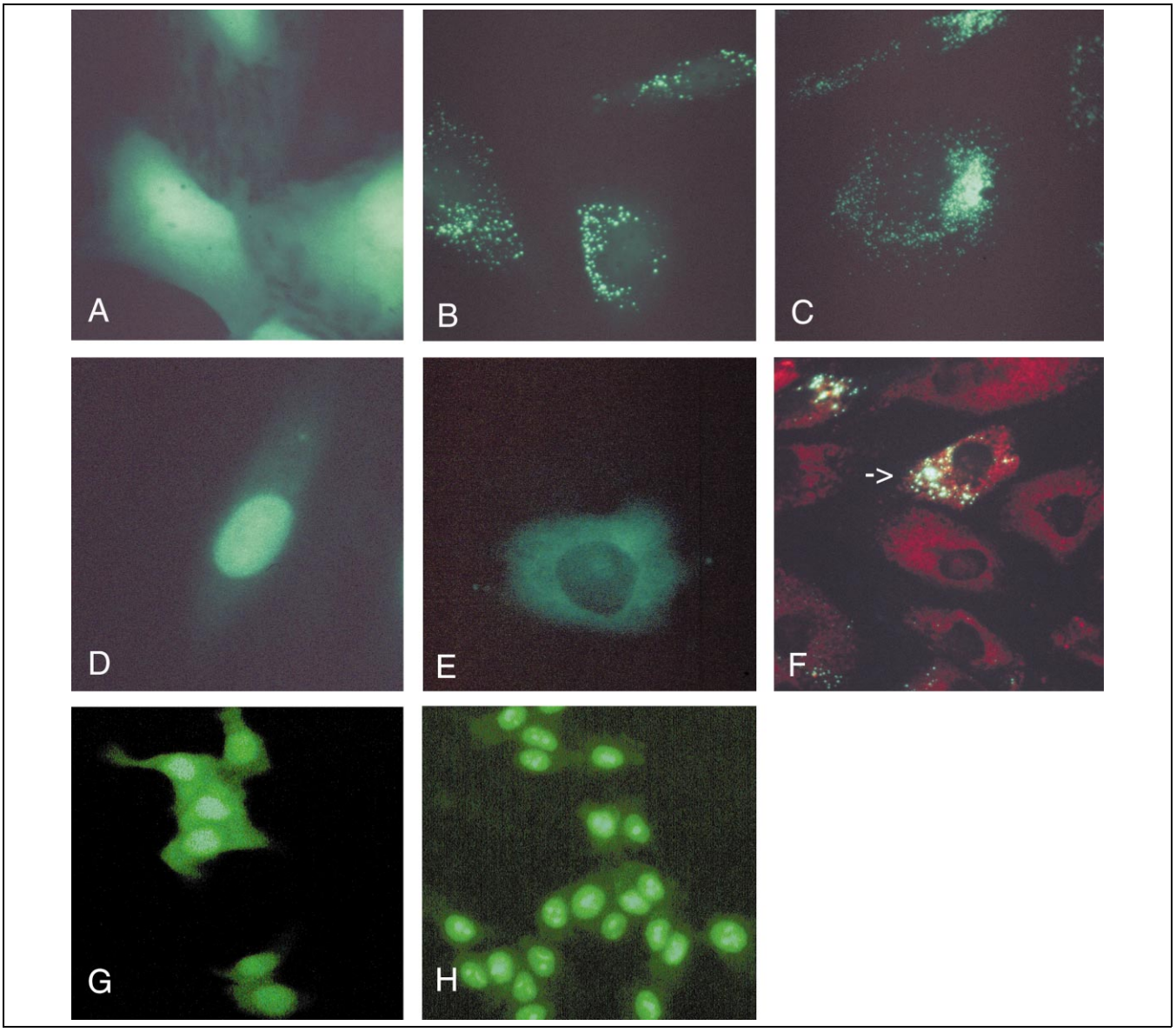


Fig. 4. Fluorescence micrographs of GFP with random peptide sequences showing different subcellular localizations. 5×10^7 HUVEC cells (A–F) were infected with a library of GFP C-terminal random 20-mers, and $\sim10^7$–10^8 A549 cells were infected with this library and loop 3 12-mer and 18-mer libraries (G,H). The EGFP scaffold is evenly distributed within the cells (A). The sequence ARIHRL causes a punctate perinuclear localization of EGFP; this pattern is similar to that observed with peroxisomes (B). The sequence SRTASTGWYTSSNLRRKSAL exhibits a different perinuclear localization, into smaller and more numerous structures than in B, also possibly peroxisomes (C). The C-terminal 20-mer KRMKTAVVRRRPLRDRRALW confers a nuclear localization to EGFP (D). The sequence CNNTSDGLRTPKKHKRGSTI causes EGFP localization in the cytoplasm (E). F: Cells expressing EGFP fused to the sequence SRTASTGWYTSSNLRRKSAL (as in C) were stained with LysoTracker (red). The red pattern, most visible in uninfected cells, gives a different pattern than the green fluorescence (arrow), suggesting that EGFP with this localization sequence is not sequestered in lysosomes. G: A typical picture of cells showing a partial nuclear localization of GFP. H: Cells showing GFP localized to substructure in the nucleus, possibly the nucleolus.

Table 2
Fluorescence microscopy screen of GFP-based peptide library localizations

	Cell clusters screened (#)	Localization phenotype (% of total cell clusters)					
		Strong nuclear	Partial nuclear	Nucleolar and nuclear	Nuclear-excluded	Punctate	Total
C-20-mer library	1800	0.45 ± 0.26	1.57 ± 0.33	0.12 ± 0.079	0.14 ± 0.10	2.54 ± 1.06	4.82
Loop 3 12-mer library	800	0.80 ± 0.29	2.16 ± 0.91	0.08 ± 0.14	0.085 ± 0.088	0.00	3.13
Loop 3 18-mer library	700	1.43 ± 0.80	2.21 ± 1.03	0.53 ± 0.44	0.12 ± 0.14	0.00	4.29
GFP alone	2000	0.00	0.00	0.00	0.00	0.00	0.00

Table 3
Sequences of peptide library members exhibiting defined subcellular localizations

Library	Peptide sequence	#	Localization
A549			
C-20-mer	PTSQL	**11**	punctate
	FMATLEARL	**12**	punctate
	RAMLKEMKM	**13**	nuclear-excluded
	GHGATYGCRRCRRVMRSLCLRPVRP	**14**	partial nuclear
	AVARAIPGDARRSRGRWRRS	**15**	partial nuclear
	IMRRRRRYEPQTPTGDTAWI	**16**	partial nuclear
	QWVKDASGARQHAARTPRIR	**17**	partial nuclear
	RPLAIRIRRLRGGSVVRRLLGP	**18**	partial nuclear
	RWSFKMSRTNSGERRRMRRH	**19**	partial nuclear
	RSAKNMPTRGSYKAPQRVSY	**20**	partial nuclear
Loop 3 12-mer	KRGLRHTRAQHR	**21**	nuclear
	PQRRPRHRKMQQ	**22**	nuclear
	RARLSRAQSTTR	**23**	partial nuclear
	KSYKVRRLGPGP	**24**	partial nuclear
	SAWSRVRRRATA	**25**	partial nuclear
	RGSKARNQSRGM	**26**	partial nuclear
	IVKSTRNRSRRG	**27**	partial nuclear
Loop 3 18-mer	YWRRGQVHRYARRRYMSG	**28**	nuclear
	SKYRTRQARQQCCLQEQS	**29**	nuclear
	VDSRRQQGRKRRHGHGSG	**30**	nuclear
	SLRTSRAVRRRHRWQIAR	**31**	nucleolar
	INRRQRRKTRRNACVGN*L*	**32**	nucleolar
	TLVPWHRNRNRRFRHPRT	**33**	nucleolar
	TVSVRGKRSASRIQTGHA	**34**	partial nuclear
	AHYHHGHTHRDKRPRAWV	**35**	partial nuclear
	RRRQCKMRQSAGLGADKT	**36**	partial nuclear
	DMHSARGHTKQTRKRYNT	**37**	partial nuclear
	MHPREEMWSRPHTRRVRP	**38**	partial nuclear
	GADSPGRMGKARTRATKR	**39**	partial nuclear
	RTGQSCRTAGRGRKGHTG	**40**	partial nuclear
	KNTESGYGLKRHKAGRVH	**41**	partial nuclear
	HRVIVGHPRRRRAPDEEK	**42**	partial nuclear
	AKSDGRYKGHHVCRGKYT	**43**	partial nuclear
HUVEC			
C-20-mer	ARIHRL	**44**	punctate
	SRTASTGWYTSSNLRRKSAL	**45**	punctate
	KRMKTAVVRRRPLRDRRALW	**46**	nuclear
	CNNTSDGLRTPKKHKRGSTI	**47**	nuclear-excluded

brary. The most infrequent patterns of localization were nucleolar and nuclear-excluded.

To verify that the peptides were responsible for these localizations, we isolated the peptide DNA by reverse transcription-polymerase chain reaction (PCR) from expressing cells, and re-introduced them into naive cells. Each subcellular localization was associated with a unique peptide sequence. Table 3 lists sequences isolated from 36 different clones. The punctate localization of Fig. 4B,C was seen in A549 cells with two additional short peptides, PTSQL and FMATLEARL. All four punctate-localized peptides were from the C-terminal library, and all had a C-terminal L. One additional sequence was nuclear-excluded, five were strongly nuclear-localized, three were nucleolar-localized, and 22 exhibited partial nuclear localization. The basic amino acid content of the punctate-localized peptides averaged 17%, the nuclear-localized peptides averaged 36%, and the nucleolar-localized peptides averaged 37%. Partial nuclear-localized peptides averaged 30% from the C-terminal 20-mer library, with a range from five to eight basic residues per peptide, and averaged 35% from the loop 3 12-mer library, where all but one sequence had four basic residues. Peptides from the loop 3 18-mer library averaged 28%, and all contained four–six basic residues.

3. Discussion

In this paper, we have examined the use of a stable protein scaffold for retroviral delivery and display of functional peptides and peptide libraries in mammalian cells. The rigid β-can structure of GFP allows placement of individual peptides into the loops tested here without no-

ticeable perturbation of the fluorophore. This allows an estimation of library member concentration in cells, an important variable for the success of phenotypic screens. Peptides as long as 18-mers were used in the random loop libraries since an average of roughly this number of residues is buried for each protein in protein–protein complexes [29] that correspond to standard-size interfaces [30]. Significantly longer random libraries were not used in the loops due to the increase in the number of library members with stop codons in the random region (these truncated GFP constructs did not fluoresce and probably did not fold [31]) and because the GFP expression level drops as the total number of inserted residues is increased. Based on FACS analysis of immunoreactive Alexa 568-stained library members, poorly fluorescent or unfolded GFP constructs are not present at significant levels within the cell.

For library sites we selected surface loops with temperature factors; such sites have a more diffuse electron density and thus a larger mean square displacement of the atoms in the crystal structure, indicating the loops are more flexible and thus that they may be more accommodating to random peptides than more rigid loops. Three loops allowed a majority of GFP constructs with random inserts to fold and fluoresce well above the background autofluorescence. Libraries based on loop 4 may have a different structural bias than loop 3 as the base of the loop is an additional 7 Å across. With only one end fused to the scaffold, peptide libraries fused to the GFP C-terminus will have a more extended bias than loop-constrained libraries. Examination of 10 different C-terminal sequences showed that these peptides are not significantly degraded in the cell. They thus appear to be suitable for phenotypic screens.

The outcome of screening for peptides that disrupt functional signaling pathways may be influenced by a number of variables, such as the structural bias of the GFP-displayed peptide libraries used compared to the bound conformational preferences of each individual target in the signaling pathway, their toxicity within the cell, the degree of constraint of a specific library member, and the complexity of the library actually screened. The steady state concentration of the library member during the time necessary to change cellular phenotype will also be important, as this will control the saturation of accessible binding partners and thus the degree of phenotypic change resulting from an individual binding partner. If the goal of a phenotypic random library screen is to dissect intracellular signaling pathways [9,10], use of a library with high expression levels may allow analysis of more peptide–target interactions than a poorly expressed library. For the single peptide characterized in [19], its solubility was in the range of 1 μM or less due to the extremely hydrophobic nature of this peptide; it may be difficult to specify the binding partners of such peptides. This poor solubility may be one way to delay or avoid intracellular proteolysis; another way may be to select for a peptide that can block or downregulate the proteasome but still give a phenotypic response. The four libraries examined here appear to be expressed at average concentrations in a range from $\sim$1 to 50 μM in A549 cells. Since these concentrations are well below those reported for the staphylococcal nuclease-based library screened in yeast [3], it was not clear how many non-uniformly localized sequences would be observable. Our results suggest that even at significantly lower average library concentrations in mammalian cells, these libraries contain peptides that can interact with cellular macromolecules. In the individual libraries there is a 100-fold to 1000-fold range of fluorescence based on the FACS profiles, thus individual sequences may be expressed at significantly different levels from the mean library concentration. The use of libraries expressed at different levels may allow a deliberate selection of peptides with different affinities that change cellular phenotype.

Peptides placed into GFP loops can retain the function of the peptide. Replacement of loop 2, 3 or 4 residues with the HA epitope tag YPYDVPDYASL allowed interaction with an anti-epitope antibody and immunoextraction of the GFP. Replacement of a loop 3 residue with two known NLSs directed the GFP to the nucleus, while replacement with control sequences, previously shown to be inactive, did not localize GFP in the nucleus. This indicates that the loop-constrained and C-terminal NLS peptides functionally interact with the nuclear pore karyopherin responsible for transport into the nucleus, and that SARs previously observed for these peptides can apply when they are presented in the context of a surface GFP loop. The SV-40 NLS peptide PPKKKRKVA is known to bind to importin α as an extended β-strand [27]. The flexible linkers at each end of loop 3, -GQGGG- and -GQAGGG-, combined with the relatively short length of this peptide, appear to allow it to adopt a more extended conformation when placed in this 20 residue loop. Shorter linkers may constrain the allowed conformations of loop peptides to those with loop/turn-like structures and may bias against extended peptides. The use of mono-glycine linkers in loop 3 appears to inactivate the SV-40 NLS and diminishes the overall expression level of the GFP scaffold. Thus the use of longer, flexible linkers may be important for both higher expression and activity of certain peptides constrained in the loop of the GFP scaffold.

The above results suggested that random loop libraries could be screened for different types of localization sequences. We examined the localization of individual library members expressed in HUVEC as well as A549 cells. Since GFP alone is uniformly distributed, defined localizations are assumed to be due to the peptide library members. Libraries with different structural biases had different frequencies of individually localized peptides. The most obvious example was the presence of punctate-appearing peptides in 2.5% of the C-terminal 20-mer library cells, but

not in the loop replacement libraries. The defined localizations resulting from the screen occurred in multiple cell types, and rescued peptides re-introduced into naive cells gave the same phenotype for 40 different peptides.

The C-terminal tripeptides of the sequences with a punctate localization were HRL, SAL, SQL and ARL. A proposed consensus sequence at the C-terminus of a protein for peroxisomal matrix-targeting in mammals is (S, A, C, K, N)(K, R, H, Q, N, S)L, and a similar punctate localization to that observed here was obtained with GFP fused to the C-terminal peptide PKSNL [32]. Two of these sequences fit this definition, and two others have two of the three consensus residues. Thus these particular peptides may localize GFP to peroxisomes, and the results suggest that the consensus definition for such localization could be expanded to include H in the first position and A in the middle position. If all punctate-localized peptides are in peroxisomes, and require a Leu at the C-terminus of the peptide, the observed frequency of 2.5% suggests that additional residues in the first two positions of the 3-mer consensus sequence will allow this localization.

Clones obtained from the localization screen contained several NLSs, two in the loop 3 12-mer library, three in the loop 3 18-mer library, and one in the C-terminal 20-mer library. Peptides **22**, **30** and **46** fit the 'classical' definition of an NLS, i.e. four consecutive basic residues (K or R), three consecutive basic residues followed by H or P, or P followed by three of four basic residues [33]. Peptide **46** also fits the definition of a bipartite NLS, i.e. two basic residues, a spacer of 10 amino acids, and then a string of three out of five basic residues. Peptides **21**, **28**, and **29**, containing four–six basic residues, appear to differ from these two types of NLS. Peptide **28** appears similar to a bipartite NLS, but with only seven intervening residues between basic regions. Peptides **21** and **29** have more evenly spaced basic residues. Thus there may be additional types of NLS beyond the classical definition [34], or these three peptides could bind to a cytoplasmic macromolecule that is independently shuttled into the nucleus.

A large number of localized peptides exhibited a partial nuclear localization, including peptides **16**, **35**, **41**, and **42**, which contain residues fitting the definition of a classical NLS. All partially nuclear peptides contained at least four basic amino acids, and a K+R content of 30%, 35% and 28% for the C-terminal, loop 3 12-mer and 18-mer libraries, respectively. These peptides were on average somewhat less basic than the NLS peptides (36% basic residue content), but they vary from 25 to 40% basic residues, suggesting that the overall basic residue content, when above a threshold, does not determine the extent of localization. If these peptides achieve nuclear localization by binding to macromolecules independently shuttled to the nucleus, such binding would appear to require a significant K+R content. Since we have found peptides with classical NLS motifs that are only partially nuclear, and fully nuclear peptides without a classical NLS motif, there may be other factors in addition to the classical NLS motifs that are important for nuclear localization. It is possible that some of the peptides saturate the transport mechanism, and thus appear partially localized due to the inability of the transport system to import all of the peptide to the nucleus.

The nuclear-excluded peptide **13**, RA**M**L**K**E**M**K**M**, contains a pattern of three out of four hydrophobic residues (bold) similar to other nuclear exclusion sequence peptides such as the HIV-1 Rev peptide **L**PP**L**ER**L**T**L**D [35], and may thus be transported by a similar CRM-1-dependent mechanism. The nuclear-excluded peptide **47** contains five basic residues in a stretch of eight amino acids and a classical NLS motif, KKXK, but does not contain the leucine or hydrophobic residue pattern typical of peptides exported by CRM-1 [36]. Its mechanism of export is thus unclear; it may bind to a macromolecule that is exported from the nucleus. The nucleolar-localized peptide **31** has seven basic residues out of 18, and is similar in sequence (bold residues) SL**R**TS**R**AV**RRR**H**RW**-QIAR to the nucleolar localization sequence (with one gap added) from the HIV Rev protein, **R**QA**R**RN**RRR**-**RW**RERQR. It is also similar to other nucleolar localization sequences in its high content of basic residues in a short sequence, suggesting that it may act by a similar mechanism.

These results illustrate the usefulness of several GFP-based peptide libraries for screening in mammalian cells. While we have found peptides that contain expected localization motifs, we have also found novel sequences that do not contain known localization motifs. This screen has examined ca. 2000 cell clusters out of three libraries of 10^7–10^8 random sequences. The frequency of peptides with an obvious phenotype, ca. one in 25, has allowed the screening of only a small part of these libraries. Using FACS analysis and different phenotypic readouts, much larger GFP-based peptide libraries (over 10^9 random sequences) have also been successfully screened in mammalian cells (E. Masuda and T. Kinsella, personal communication). The utility of this approach may be enhanced by the addition of fixed localization sequences, such as the ones examined here, to direct libraries to defined cellular locations, as well as the use of libraries with other structural biases, such as to α-helical or β-strand peptides, that may be under-represented in the libraries examined here.

4. Significance

The combination of in vitro combinatorial chemistry with highly efficient screens has become an important methodology in pharmaceutical chemistry. The production of intracellular peptide libraries in mammalian cells, at concentrations high enough to allow their use in phenotypic screens, would be an analogous process. Peptides

which successfully change cellular phenotype, and which thus may have a functional role in the development of the phenotype, could have many uses. They could be used to help define the signaling pathway by identifying one or more key members coupled to the phenotypic change. This could be done by affinity-extracting the peptide binding partners, which could be involved in their mechanism of action, and identifying them using highly sensitive mass spectrometry techniques. Some of the binding partners could be of interest to pharmaceutical chemists as potential therapeutic targets. Peptides resulting from a phenotypic screen could also represent a starting point for the development of peptidomimetic drugs that act on the signaling pathway. If they are or could be engineered to be stable within the cell, they could be useful for gene therapy of defined diseases. Discovery of different peptide sequences with a varying function could allow the rapid definition of critical residues and the construction of in vivo SARs. Differential use of these screens in normal cells versus cells derived from diseased organisms could help elucidate changes in signaling pathways which accompany disease, giving a more detailed understanding of disease processes.

5. Materials and methods

5.1. Materials

Anti-*Aequorea victoria* GFP rabbit serum, Alexa 568-conjugated goat anti-rabbit IgG, and LysoTracker were obtained from Molecular Probes (Eugene, OR, USA), goat anti-rabbit IgG–horseradish peroxidase conjugate was from Sigma (St. Louis, MO, USA), rabbit polyclonal anti-flag probe was from Santa Cruz Biotechnology (Santa Cruz, CA, USA), anti-HA affinity matrix was from Babco (Berkeley, CA, USA), EGFP and recombinant EGFP were from Clontech (Palo Alto, CA, USA), and complete EDTA-free protease inhibitor cocktail was from Boehringer Mannheim (Chicago, IL, USA). All restriction endonucleases were from New England Biolabs (Beverly, MA, USA). All cell growth media were from JRH Biosciences (Williamsburg, VA, USA) except as noted.

5.2. Cells

Phoenix E and Phoenix A 293T retroviral packaging cell lines [37] were carried in 10% fetal bovine serum with 1% penicillin–streptomycin and Dulbecco's modified Eagle media (Mediatech Cellgro, Herndon, VA, USA). A549 human lung adenocarcinoma cells were carried in 10% fetal calf serum, and 1% penicillin–streptomycin in F12K Kaighn's modified media. Jurkat E T cell leukemic cells stably expressing the ecotropic receptor were carried in 10% fetal calf serum with 1% penicillin–streptomycin in RPMI 1640 media. Calcium phosphate transfection of Phoenix E and Phoenix A cells and infection of A549 and Jurkat E cells were carried out as described [37].

5.3. Vectors

Except as noted, all retroviral constructs were based on p96.7, a retroviral vector with a composite CMV promoter fused to the transcriptional start site of the MMLV R-U5 region of the LTR, an extended packaging sequence, deletion of the MMLV gag start ATG, and a multiple cloning region. This vector is identical to pCGFP [38] without the EGFP (Clontech) cassette. To create the vectors used in Fig. 1, coding regions for EGFPs modified with linker-HA epitope-linker G_4YPYDVPDYASLG_4 test peptides in loops 2, 3, and 4 were removed by PCR from vectors obtained from Randall Armstrong (Rigel Inc.) and put into 96.7. The EGFP vectors used in Figs. 2–4 were based on 96.7 EGFP, a retrovirus expression vector containing a Kozak consensus start [39] followed by a modified EGFP coding region which has 21 silent mutations introducing 12 restriction sites into the open reading frame of Clontech's EGFP gene (results in eight non-optimized codons). The library vectors 96.7E3Z and 96.7E4Z were derived from p96.7EGFP by replacing the sequences corresponding to residues of loop 3 (D134) and of loop 4 (G190–D191) with a sequence encoding the *Bst*XI-encoding flexible linkers and a stuffer fragment (-GQGGG-stuffer-GQAGGG-). Vectors, used in Fig. 3, coding modified GFPs with NLS and control peptides flanked with the flexible linkers in loop 3 were made by ligating annealed oligos encoding the peptide inserts (Flag: 5′-GTGGCGTCGCCGACTACAAGGACGACGACGACAAGG-CAGTAGGCCAAGCAG and 5′-TTGGCCTACTGCCTTGTC-GTCGTCGTCCTTGTAGTCGGCGACGCCACCGCC, SV-40: 5′-GTGGCCCTCCAAAAAAGAAGAGAAAGGTAGCTGGC-CAAGCA and 5′-TTGGCCAGCTACCTTTCTCTTCTTTTTT-GGAGGGCCACCGCC, SV-40mut: 5′-GTGGCCCTCCAAA-GACCAAGAGAAAGGTAGCTGGCCAAGCA and 5′-TTGG-CCAGCTACCTTTCTCTTGGTCTTTGGAGGGCCACCGCC, c-myc: 5′-GTGGCCCTGCCGCCAAAAGGGTGAAGCTGG-ACGGCCAAGCAG and 5′-TTGGCCGTCCAGCTTCACCCT-TTTGGCGGCAGGGCCACCGCC, c-mycmut: 5′-GTGGCC-CTGCCGCCAAAACCAAGAGGCTGGACGGCCAAGCAG and 5′-TTGGCCGTCCAGCCTCTTGGTTTTGGCGGCAGG-GCCACCGCC) into the *Bst*XI sites of p96.7E3Z. The short linker SV-40 NLS peptide was added to 96.7 EGFP by PCR sewing together fragments created with a GFP 5′ primer plus 5′-AGCTACCTTTCTCTTCTTTTTTGGAGGGCCCTCGATGT-TGTGGCGGATCTTG, and a GFP 3′ primer plus 5′-CCT-CCAAAAAAGAAGAGAAAGGTAGCTGGCTCCGTGCAG-CTCGCCGACCACTAC. The vector expressing an EGFP with C-terminally fused SV-40 NLS peptide and rigid -EEAAKA-linker was constructed by ligating annealed oligos (cSV-40NLS: 5′-TCGAGTTCGTGACCGCCGCCGGGATCACTCTCGGC-ATGGACGAGCTGTACAAGGGCCCTCCAAAAAAGAAG-AGAAAGGTAGCTTAGC and 5′-GGCCGCTAAGCTACCT-TTCTCTTCTTTTTTGGAGGGCCCTTGTACAGCTCGTCC-ATGCCGAGAGTGATCCCGGCGGCGGTCACGAAC) into the *Xho*I/*Not*I sites of p96.7EGFP, which also introduces flanking *Sfi*I sites. The p96.7EZ vector was made by ligating a *Sfi*-digested PCR fragment into the *Sfi*I sites of the previous vector. For the creation of the libraries used in the phenotypic screen, the ex-

pression cassettes from 96.7E3Z, 96.7E4Z, and 96.7EZ were moved into the *Eco*RI/*Not*I sites of TRA-Beau, a TRA retrovirus vector [38] containing a slightly altered multiple cloning site.

5.4. Peptide libraries

For the EGFP loop 3 and 4 libraries in TRA, small test libraries of ca. 2×10^7 primary transformants were constructed. Degenerate oligonucleotide pools encoding the random 12- or 18-mer peptide libraries 5′-ATCGATCTCCACCTGCTTGG-CC$(MNN)_{12/18}$GCCACCGCCCTGGCCAGGCTC synthesized with an NNK codon bias in order to reduce the stop codon frequency, were annealed to a primer 5′-GAGCGAGCCTGGC-CAGGGCGGTGGC and extended using the Expand PCR kit (Roche Molecular Biochemicals, Germany). The double-stranded library insert was digested with *Bst*XI, ligated into the *Bst*XI sites of TRAE3Z and TRAE4Z, and electroporated into *E. coli* for amplification on plated media. The C-terminal 20-mer library was created in a similar manner with the primer 5′-GATCGACG-TGGCCGCCAAGGCC and library template 5′-AGTATGG-CCTCACGGGCCTA$(MNN)_{20}$GGCCTTGGCGGCCACGTC-GA, digested with *Sfi*I and put into the *Sfi*I sites of TRAEZ. The total plated complexity of greater than 2×10^9 transformants was divided into fractions of $\sim2\times10^8$ for screening purposes.

5.5. FACS

Flow cytometric analysis of GFP expression in all cells was performed on a FACScan (Beckton-Dickinson, San Jose, CA, USA) unless otherwise noted, and analyzed using FloJo software (Treestar Software, San Carlos, CA, USA).

5.6. Localization studies

Retroviral vectors containing wild type EGFP or the C-terminal GFP library ($\sim2\times10^8$ random sequences) were introduced into HUVEC cells using the 96.7 vector, and were observed 4 days after infection. GFP or GFP peptide libraries ($\sim2\times10^7$ random sequences in the loop 3 libraries, 2×10^8 random sequences in the C-terminal library) were introduced into A549 cells using the TRA vector [38] and localization was also observed after 4 days. Localization was determined by direct examination of live cells growing on 10 cm culture dishes using a Nikon Elipse TE300 fluorescence microscope. Independent examinations of localization by three experimenters were averaged.

5.7. Immunoprecipitations, SDS–PAGE and Western blotting

For the preparation of whole cell lysates, cells were collected, washed and lysed by freeze-thaw/vortexing in lysis buffer (50 mM HEPES pH 7.4, 150 mM NaCl, 5 mM EDTA, 5 mM EGTA, 1% Triton X-100) with added Complete Protease Inhibitor cocktail (Boehringer Mannheim, Chicago, IL, USA). Lysate cleared by centrifugation was boiled in 4× Nupage (Novex, San Diego, CA, USA) sample buffer and loaded onto SDS gels as per the manufacturer's recommendation. For immunoprecipitations, 7.5×10^5 cells were collected 4 days post-infection, washed and lysed by freeze-thaw in 100 μl 1× PBS with Complete Protease Inhibitor cocktail. 15 μl anti-HA affinity matrix (Babco, Berkeley, CA, USA) was added to cleared lysate and incubated for 14 h at 4°C. Beads were washed with PBS and boiled in 4× Nupage sample buffer. SDS–PAGE was performed on 4–12% Nupage Bis-Tris Polyacrylamide Gels with MES running buffer according to the manufacturer's recommendations (Novex, San Diego, CA, USA). Samples were transferred to polyvinylidene difluoride membranes and blocked using 10% milk, 0.1% Tween 20 in 1× PBS. Antibodies were diluted into the same blocking buffer. Primaries were used at 1:2000, secondaries at 1:5000. Membranes were developed using an ECL plus enhanced chemiluminescence kit (Amersham Pharmacia, Piscataway, NJ, USA) and detected using Hyperfilm ECL (Amersham Life Science, Buckinghamshire, UK).

5.8. MALDI-TOF mass spectrometry

GFP-fused random peptides or peptide fragments from a lys-C protease digest were examined on a Bruker Reflex III mass spectrometer after affinity purification using an immobilized anti-GFP antibody (Molecular Probes, Eugene, OR, USA) fused to protein A/G beads. The GFP fusions were examined in linear mode using sinapinic acid as a matrix. A saturated sinapinic acid solution in 30% acetonitrile–0.1% trifluoroacetic acid was used to elute the beads, and 0.5 μl was dried onto the sample plate. Spectra were externally calibrated with bovine serum albumin and myoglobin. Lys-C-digested peptides, produced from GFP fused to different random sequences under non-denaturing conditions, were examined in reflectron mode using as a matrix α-cyano 4-hydroxy-*trans*-cinnamic acid dissolved in 30% acetonitrile–0.1% trifluoroacetic acid. Use of native conditions focused proteolysis on the C-terminal peptide fused to GFP. Digestion was in 50% ^{18}O-H_2O–50% ^{16}O-H_2O so that the most C-terminal peptide was the only one without dual ^{16}O–^{18}O incorporation and a doublet parent ion. ACTH and angiotensin were used as standards for internal calibration.

5.9. Estimation of the cellular concentration of GFP libraries

The intracellular concentration of each library was estimated using the measured absorption at 489 nm of a fixed number of lysed cells, the Beer–Lambert law and a molar extinction coefficient at 489 nm for GFP of 55 000 M^{-1} cm^{-1}. The volume of a single cell (2.5 pl for an A549 cell) was estimated using the cell's diameter obtained from a hemocytometer, assuming the cell is spherical. 3×10^5 cells were used for this measurement.

References

[1] G. Caponigro, M.R. Abedi, A.P. Hurlburt, A. Maxfield, W. Judd, A. Kamb, Transdominant genetic analysis of a growth control pathway, Proc. Natl. Acad. Sci. USA 95 (13) (1998) 7508–7513.
[2] C. Geyer, A. Colman-Lerner, R. Brent, 'Mutagenesis' by peptide

aptamers identifies genetic network members and pathway connections, Proc. Natl. Acad. Sci. USA 96 (1999) 8567–8572.
[3] T. Norman, et al. A. Murray, Genetic selection of peptide inhibitors of biological pathways, Science 285 (1999) 591–595.
[4] W. Hilt, D. Wolf, Proteasomes: destruction as a programme, Trends Biochem. Sci. 21 (1996) 96–102.
[5] A. Goldberg, T. Akopian, A. Kisselev, D. Lee, M. Rohrwild, New insights into the mechanisms and importance of the proteasome in intracellular protein degradation, Biol. Chem. 378 (1997) 131–140.
[6] D. Bromme, A. Rossi, S. Smeekens, D. Anderson, D. Payan, Human bleomycin hydrolase: molecular cloning, sequencing, functional expression, and enzymatic characterization, Biochemistry 28 (1996) 6706–6714.
[7] J. Beninga, K. Rock, A. Goldberg, Interferon-γ can stimulate post-proteasomal trimming of the N terminus of an antigenic peptide by inducing leucine aminopeptidase, J. Biol. Chem. 273 (1998) 18734–18742.
[8] A. Hattori, H. Matsumoto, S. Mizutani, M. Tsujimoto, Molecular cloning of adipocyte-derived leucine aminopeptidase highly related to placental leucine aminopeptidase/oxytocinase, J. Biochem. (Tokyo) 125 (1999) 931–938.
[9] S. Wilk, E. Wilk, R.P. Magnusson, Purification, characterization, and cloning of a cytosolic aspartyl aminopeptidase, J. Biol. Chem. 273 (1998) 15961–15970.
[10] M. Belich, J. Trowsdale, Proteasome and class I antigen processing and presentation, Mol. Biol. Rep. 21 (1995) 53–56.
[11] D.H. Lee, A.L. Goldberg, Proteasome inhibitors: valuable new tools for cell biologists, Trends Cell Biol. 8 (1998) 397–402.
[12] G. Siligardi, A.F. Drake, The importance of extended conformations and, in particular, the PII conformation for the molecular recognition of peptides, Biopolymers (Pept. Sci.) 37 (1995) 281–292.
[13] R. Hodges, De novo design of α-helical proteins: basic research to medical applications, Biochem. Cell Biol. 74 (1996) 133–154.
[14] S. Futaki, Peptide ion channels: design and creation of function, Biopolymers 47 (1998) 75–81.
[15] K. Beck, B. Brodsky, Supercoiled protein motifs: the collagen triple-helix and the α-helical coiled coil, J. Struct. Biol. 122 (1998) 17–29.
[16] P. Colas, B. Cohen, T. Jessen, I. Grishina, J. McCoy, R. Brent, Genetic selection of peptide aptamers that recognize and inhibit cyclin-dependent kinase 2, Nature 380 (1996) 548–550.
[17] M.R. Abedi, G. Caponigro, A. Kamb, Green fluorescent protein as a scaffold for intracellular presentation of peptides, Nucleic Acids Res. 26 (1998) 623–630.
[18] S.D. Nuttall, M.J. Rousch, R.A. Irving, S.E. Hufton, H.R. Hoogenboom, P.J. Hudson, Design and expression of soluble CTLA-4 variable domain as a scaffold for the display of functional polypeptides, Proteins 36 (1999) 217–227.
[19] X. Xu, C. Leo, L. Cindy, Y. Jang, E. Chan, D. Padilla, B. Huang, T. Lin, T. Gururaja, Y. Hitoshi, J. Lorens, D. Anderson, B. Sikic, Y. Luo, G. Nolan, D. Payan, Dominant effector genetics in mammalian cells, Nat. Genet. 27 (2001) 23–29.
[20] F. Yang, L. Moss, G. Phillips, The molecular structure of green fluorescent protein, Nat. Biotechnol. 14 (1996) 1246–1251.
[21] M. Ormo, A. Cubitt, K. Kallio, L. Gross, R. Tsien, S. Remington, Crystal structure of the *Aequorea victoria* green fluorescent protein, Science 273 (1996) 1392–1395.
[22] G.N. Phillips, Structure and dynamics of green fluorescent protein, Curr. Opin. Struct. Biol. 7 (1997) 821–827.
[23] A. Radu, G. Blobel, M.S. Moore, Identification of a protein complex that is required for nuclear protein import and mediates docking of import substrate to distinct nucleoporins, Proc. Natl. Acad. Sci. USA 92 (1995) 1769–1773.
[24] J. Moroianu, G. Blobel, A. Radu, Previously identified protein of uncertain function is karyopherin α and together with karyopherin β docks import substrate at nuclear pore complexes, Proc. Natl. Acad. Sci. USA 92 (1995) 2008–2011.
[25] D. Kalderon, B. Roberts, W. Richardson, A. Smith, A short amino acid sequence able to specify nuclear location, Cell 39 (1984) 499–509.
[26] E. Conti, M. Uy, L. Leighton, G. Blobel, J. Kuriyan, Crystallographic analysis of the recognition of a nuclear localization signal by the nuclear import factor karyopherin α, Cell 94 (1998) 193–204.
[27] C. Dang, W. Lee, Identification of the human c-myc protein nuclear translocation signal, Mol. Cell. Biol. 8 (1988) 4048–4054.
[28] J. Makkerh, C. Dingwall, R. Laskey, Comparative mutagenesis of nuclear localization signals reveals the importance of neutral and acidic amino acids, Curr. Biol. 6 (1996) 1025–1027.
[29] J. Janin, C. Chothia, The structure of protein–protein recognition sites, J. Biol. Chem. 265 (1990) 16027–16030.
[30] L. Lo Conte, C. Chothia, J. Janin, The atomic structure of protein–protein recognition sites, J. Mol. Biol. 285 (1998) 2177–2198.
[31] J. Dopf, T. Horiagan, Deletion mapping of the *Aequorea victoria* green fluorescent protein, Gene 173 (1996) 39–44.
[32] L. Amery, C. Brees, M. Baes, C. Setoyama, R. Miura, G. Mannaerts, P. van Veldhoven, C-terminal tripeptide Ser-Asn-Leu (SNL) of human D-aspartate oxidase is a functional peroxisome-targeting signal, Biochem. J. 336 (1998) 367–371.
[33] G. Hicks, N. Raikhel, Protein import into the nucleus: an integrated view, Annu. Rev. Cell. Dev. Biol. 11 (1995) 155–188.
[34] D. Christophe, C. Christophe-Hobertus, B. Pichon, Nuclear targeting of proteins: how many different signals?, Cell Signal. 12 (2000) 337–341.
[35] O. Rosorious, B. Fries, R. Stauber, N. Hirschmann, D. Bevec, J. Hauber, Human ribosomal protein L5 contains defined nuclear localization and export signals, J. Biol. Chem. 275 (2000) 12061–12068.
[36] B. Henderson, A. Eleftheriou, Comparison of the activity, sequence specificity, and CRM1-dependence of different nuclear export signals, Exp. Cell. Res. 256 (2000) 213–224.
[37] S. Swift, J. Lorens, P. Achacoso, G. Nolan, Rapid production of retroviruses for efficient gene delivery to mammalian cells using 293T cell-based systems, in: J. Coligan, A. Kruisbeek, D. Margulies, E. Shevach, W. Strober (Eds.), Current Protocols in Immunology, Vol. 10.17C, John Wiley and Sons, Inc., New York, 1999, pp. 1–17.
[38] J. Lorens, M. Bennett, D. Pearsall, W. Throndset, A. Rossi, R. Armstrong, B. Fox, E. Chan, Y. Luo, E. Masuda, D. Ferrick, D. Anderson, G. Nolan, D. Payan, Retroviral delivery of peptide-modulators of cellular functions, Mol. Ther. 1 (5) (2000) 438–447.
[39] M. Kozak, An analysis of 5′-noncoding sequences from 699 vertebrate messenger RNAs, Nucleic Acids Res. 15 (1987) 8125–8148.